

2010 CCPRC Annual Meeting
Workshop Session A-2
October 21, 2010, 10:45 a.m.

New Research on QRIS: Implications for Policy and System Improvement

Description

This breakout session provided an opportunity to discuss emerging research on QRIS covering a range of issues including quality measurement, approaches to validation of QRIS rating tools, quality improvement approaches, parent awareness of QRIS, and the role of QRIS in building systemic supports for quality.

Facilitator

James Elicker, Purdue University

Presenters

Gretchen Kirby, Mathematica Policy Research
Kathryn Tout, Child Trends
Lizabeth Malone, Mathematica Policy Research
Michel Lahti, University of Southern Maine
Gail Zellman, RAND Corporation

Scribe

Laura Rothenberg, Child Trends

1. Documents in Session Folder

- “Reflections on QRIS Evaluation: Quality for ME,” Michel Lahti
- “Cross-System Evaluation of QRS: Research Questions and Approaches in the QRS Assessment;” Gretchen Kirby, Kathryn Tout, and Liza Malone

2. Summary of Presentations

- **Summary of Presentation #1:** Gretchen Kirby and Kathryn Tout
 - The goals of the QRS assessment project are to catalog comprehensive and consistent information on QRIS key elements; gather and synthesize in-depth information on planning, design, and evaluation for select QRIS; analyze secondary data to aggregate definitions across QRIS; and develop a toolkit for QRIS evaluation.
 - The Compendium was the foundational work for the QRS Assessment. It made a unique contribution in that it provided a common language and rubric for describing QRIS standards and elements.
 - As a cross-system evaluation, the aim of the case study and the secondary analysis components of the Assessment are to focus on some of the key research questions coming out of current knowledge and the Compendium. The purpose is to investigate the variance that exists across QRIS and the reasons for variation. They also wanted to identify data availability and quality.
 - Two in-depth studies are underway.

- The first is “Quality Measurement in Select QRIS,” which involves looking at five select QRISs to understand what the standards look like, what quality means, and how it is defined in each system (for example, how is level one in one QRIS compared to level one in another QRIS?). The study will examine the processes used to measure the individual components that go into a quality rating (who does this, how is it documented, and how reliable is it, etc.).
 - The second study is “QRIS as a System Builder,” which explores how and to what extent QRISs contribute to ECE system development and integration (what is the role of QRIS and how might it look different in other States?). The study will include an in-depth examination of two QRIS. The study will also examine the indicators that can be used as measures in system development.
- The secondary data analysis component of the Assessment is challenging and interesting. The QRISs included in the secondary data analysis are a subset of the QRISs in the quality measurement in-depth study (three of five).
- Data sources included in the secondary analysis are QRIS databases, PD registries, and ERS data. The analysis will build on the 13 quality component categories described in the Compendium. The intent is to examine how providers differ on quality measurement (by component) along a common metric across QRISs. Also being examined are correlations among components, between observed quality and other components, and between all components and the rating level. Some issues arise when determining a common metric to use in the analysis across the three QRISs and then determining what that will mean to QRIS.
- Kathryn highlighted some of the findings and big picture pieces from the Compendium.
 - It was a first chance to get State reports of how many eligible participants are in QRIS. Across different bands of participation, they found that most States have 30% or fewer providers participating, and on average, 20% participation. States with around 60% participation tended to be States with a mandatory component. This raised interesting questions regarding capacity, as well as questions surrounding which programs were in the 20%.
 - Quality improvement was not addressed deeply in the compendium, but others have done this (National Center for Children in Poverty).
 - There were also some interesting measurement mechanics issues, including two ways of measuring QRIS: (1) a building block system where you need all indicators at one level before you can be rated at the next level; or (2) a system where you achieve points for meeting certain indicators. Point system structures tend to result in higher percentages of providers at higher ratings.
 - There are some elements such as professional development that are included in all 26 QRISs. However other indicators, such as cultural sensitivity, are included in far fewer systems.
 - Much more work is still needed on measurement.
- Following Kathryn’s presentation, there were questions on the following topics:
 - *Have you explored if QRIS systems included regulations?* There is a link to regulations, but it is not directly included.
 - *In selecting the five States for the in-depth study, will you select States that have child outcome data?* The original intention was to use States that had child

outcome data, but there weren't enough States with such data. The main criterion in choosing the five States was to select ones that used similar rating systems.

- **Summary of Presentation #2: Michel Lahti**
 - Michel explained that the aim of their study was to focus on quality standards. They have looked at both center-based and family child care. Along with other data, they collected data from parents through an anonymous survey and from staff as well.
 - Beginning in 2004, concept mapping was used to show similarities and differences between parents, stakeholders and providers. The idea of performance management and continuous improvement has been consistent throughout the design and implementation of the system.
 - From the database of programs enrolled, programs are randomly selected for evaluation. They are linked using administrative data, such as licensing data, etc. The State agency checks with providers to see if there is evidence that standards are being met. This is separate from the evaluation work done as part of the study. Data collected is brought to a team each week to be analyzed and discussed. The intent is to look at this data together with research and program staff throughout the process. This has been helpful because it is consistent with the goal of continuous improvement.
 - Connected with capacity building, Michel has been exploring how the performance management system is considered. They have conversations about how to convey the information to stakeholders, etc. The State agency has started to use the information collected in decision-making.
 - They want to continue to emphasize the importance of maintaining both criterion and translation validity. Their team includes people with very different backgrounds and interests, so they have attempted to create a structure to ensure all aspects of the project are captured. Part of the work going forward this year with INQUIRE is to look at lessons learned in trying to validate systems.
 - After Michel's presentation, the following issues were raised:
 - It may be useful to introduce a measurement based specifically on QRIS. There needs to be validity to what you are using, but it also needs to be appropriate.
 - There needs to be a match between ERS assessments, but it is also important to capture whether or not a provider has settled at a level or not.
- **Summary of Presentation #3: Gail Zellman (Reflections)**
 - Gail commended efforts that examine validity, which she explained are very important but so rarely explored. She said that not validating these measures and ratings now may complicate much that comes with rating in the future. It is important to understand that validation refers to a process that involves the use of specific measures in specific contexts. This is problematic because most of the measures used in QRISs around the country were developed for a different process (and not high-stakes settings). In many State QRISs, the data suggest many problems. It is very important that State administrators begin to understand how important this is and begin validating the measures used.
 - Validation involves looking at the pattern of relationships among measures of the same concept (e.g., if you're looking at a measure of environmental quality, other

- measures of general environmental quality); looking at items within the measure (which should relate but not duplicate); getting expert advice on how to capture the concept within the measure; and looking at the pattern of relationships between the measures you are trying to measure and other variables that should be related.
- Validation is a big deal. Work in Colorado was stopped because some of the component measures were not working. They wanted a measure of parent involvement, but did not have one that was usable. This needs to be explored in future research.

3. Summary of Discussion with Presenters and Participants.

- *Would it be important to explain to States what information they can get for secondary analysis?* Gretchen: The in-depth study is looking at how components in different systems are being measured. They are looking at developing variables across these 13 categories and loading them in a way that captures when a category doesn't work in a certain State or system. With administrative data, there is not an exogenous measure of quality to evaluate against to determine if there is a meaningful difference in the components.
- *What are the five States?* Right now, Illinois, Florida Miami-Dade, and Tennessee are included.
- *Who are the kids in CCDF and what settings are they are in? To what extent will data from the chosen States help us respond to the important questions?*
 - Measures that are used may need to vary in a non-trivial way across settings. For example, the ways in which parents engage family child care are very different than in centers, and if you are trying to come up with a measure that will work across settings, then you have a real challenge. This raises questions as to whether a rating in a center should be considered in the same way as a rating for a home.
 - Jim Elicker: In Indiana, we are looking at types of care by family SES. We will be able to get information as to whether kids from lower SES are gaining access to higher quality care.
 - Gretchen: for purposes of the analyses, we are focused on centers because we have the ERS scores for them. Illinois has ratings for license-exempt providers, so that will be captured in the in-depth study.
 - Kathryn: one of the things we're struggling with in our sample in Minnesota is that we have a sample of children in mostly three-star and four-star arrangements, and very few in lower levels. We may be making some of these decisions based on a very select group of programs, so we need to be careful about the conclusions we come to when we know we have a sample that might be weird, frankly. It is a voluntary system, so that creates a certain type of sample.
 - In Indiana, participation rates are very dynamic.
 - Ivelisse: States that were chosen were driven by the questions that we needed to answer and the data that was available.
 - QRIS is about accountability and program improvement. The whole issue of validation is complicated, and there are multiple pieces of information as to whether or not it is valid. However, you always have to start with the goals of the program. QRIS evaluation does not have to tell us everything about the programs.

- The fact that we want parents to use the ratings in making child care decisions is a very different purpose than setting standards and performance management.
- Michel: There are frameworks and ideas in other sectors to do this. At least we are having a conversation so that developing these systems brings us somewhere between licensing and accreditation. This differs depending on the stakes, but this is still a very important conversation to have.
- Gail: In the RAND study, hospitals highlighted how unfair comparisons were because some hospitals may have sicker patients, etc. In translating this to child care research, we have an obligation to be thinking about these possibilities. We do not want to penalize providers who face greater challenges.
- *What is the distinction between QRS and QRIS?* They are used interchangeably.
- On the ERS, it seems many States that use measurements of classroom quality need to move to the location level. We need to define what a classroom is.
 - Gail: In the RAND study, we have been trying to assess ways to explore variation across locations and within locations.
 - Michel: We use the ERS to average classrooms within a building. However, there are many issues surrounding this. Right now, we are using the mean score per program, where we look at a third of the classrooms. That is what we are doing now, and we may find out something different later.
- *Are there plans to make the Compendium dynamic?* Perhaps, it could be available online. Ivelisse: Our intention is to keep it updated. Many States have already made changes since we made it.
- *Have you looked at the degree to which parents make decisions about child care based upon QRIS?* Kathryn: Very few of the 26 sites we examined have any money for marketing. This is a huge issue. We have been doing some initial work in our study in Minnesota to ask if parents have heard of Parent Aware, and very few of them have.
- Gretchen: Because there are so many outstanding issues about what the stars mean, QRS is shifting toward being a supply side mechanism.